The Bayesian Causal Inference of Body Ownership Model: Use in VR and Plausible Parameter Choices

Moritz Schubert* Dominik Endres†

Philipps University of Marburg

ABSTRACT

Experiencing virtual body ownership is an important component of user experience in virtual reality applications with embodied avatars. A functioning model of body ownership could allow designers of such applications to predict the occurrence of body ownership illusions in users. One attempt at such a model, the Bayesian Causal Inference of Body Ownership (BCIBO) model, explains body ownership as inference about the causes of sensory signals. When sensory signals under consideration (e.g. tactile and visual signals) are attributed to a single object (e.g. a rubber hand), then this object is incorporated into the body. We investigate an unrealistic choice of parameter values in the original specification of the BCIBO model and make some suggestions for improvements.

Index Terms: Mathematics of computing—Probability and statistics—Probabilistic inference problems—Bayesian computation; Applied computing—Law, social and behavioral sciences—Psychology

1 INTRODUCTION

In many virtual reality (VR) applications, the user is situated in the virtual environment via an avatar. In order for the user to be able to focus on the task at hand, a sense of body ownership (SoBO) over said avatar is crucial. We define SoBO as the experience of a body as one's own. A lack of SoBO would likely lead to a feeling of discomfort (i.e. not feeling comfortable in one's "virtual skin") and impede the overall user experience.

A well-founded understanding of the mechanisms underlying the occurrence of SoBO can help VR application designers in creating appealing software for their customers. To this end, creating computational models of SoBO is a promising approach, because – given their assumptions are correct – they facilitate the prediction of changes in the modeled outcome based on the inputs and parameters included in the model.

We hold that the most useful model of this kind would be an approximation of the data generating function in the real world, i.e. a generative model of how internal (e.g. neural activity) and external (e.g. sensory input) factors cause the SoBO in humans. Our reasons for this stance are twofold: For one, this goal is very much in line with the general project of SoBO research, because in computational terms said research is nothing but an attempt to learn about the data generating function. Secondly, out of all the possible descriptions of SoBO the data generating function generalizes to a much larger number of situations than a situation-specific classifier or similar data-driven approaches. Accordingly, a good approximation of the generating function should avoid both under- and overfitting and can be helpful in a wide variety of applications.

One attempt at finding such a generative model which has been gaining traction recently is the Bayesian Causal Inference of Body Ownership (BCIBO) model [13]. Once further developed, it should be easily applicable to seated VR problems in the real world, because it assumes a seated user that is changing the position of their arms, but not the overall position of their body.

In this position paper we are going to give a brief overview of the BCIBO model (Sect. 2.1), point out a relevant flaw in its assumptions (Sect. 2.2), give suggestions for correcting said flaw (Sect. 3) and offer an outlook on possible future applications for a revised model (Sect. 4).

2 THEORY

Perhaps the most widespread paradigm to study body ownership is the rubber hand illusion (RHI) [1]. During a typical RHI experiment the participant is seated with one of their arms resting on a table in front of them. A rubber hand is placed on the table in an anatomically plausible position. The real hand is hidden by a screen and the shoulder covered by a blanket, out of which the rubber hand protrudes. Therefore, at first sight it might look to the participant like the hand in front of them is their real hand.

Rubber hand and real hand are stroked simultaneously by the experimenter with a brush. This often results in *touch referral* [17], i.e. feeling the touch of the brush on the rubber hand instead of the real one. Most of the time, touch referral is accompanied by a SoBO towards the rubber hand [1, 10].

2.1 The BCIBO Model

Samad, Chung, and Shams [13] explain the RHI with the BCIBO model. Roughly speaking, Bayesian inference is a mathematically rigorous method of updating current knowledge in light of new observations. In cognitive science and psychology it is often used to model ideal observers, i.e. agents that make the best possible use of sensory information [3, 9]. This is also how it is used in the BCIBO model. In the following we will briefly give a more technical definition of Bayesian inference.

The variable H (for hypothesis) will refer to our knowledge about the world and D will refer to the newly observed data. The Bayesian framework represents the uncertainty inherent in our knowledge about the world in the form of probability distributions over random variables, which are the quantities of interest. Inference is accomplished through the use of Bayes' theorem: $p(H|D) = \frac{p(D|H)p(H)}{p(D)}$, where $p(H)$, the *prior* distribution, represents our knowledge of the world before seeing any of the data, $p(D|H)$, the *likelihood*, represents the conditional probability of the data under our different hypotheses, and $p(D)$, the *marginal likelihood*, represents the probability of our data, irrespective of the value of any of the hypotheses under consideration. By multiplying the prior times the likelihood and normalizing it by the marginal likelihood, we arrive at $p(H|D)$, the *posterior*, which represents our knowledge about the world updated by the data.

Bayesian *causal* inference applies Bayes' theorem to the search for the causes of events (such as sensory input) [9]. The BCIBO model applies Bayesian causal inference to the SoBO. The paradigm has also been used to model other psychological phenomena, including multisensory integration in stimulus localization [9] and speech perception [12].

*e-mail: moritz.schubert@uni-marburg.de
†e-mail: dominik.endres@uni-marburg.de

978-0-7381-1367-8/21/$31.00 ©2021 IEEE
DOI 10.1109/VRW52623.2021.00019

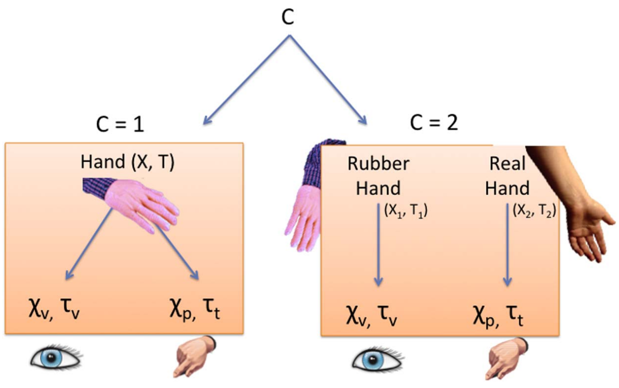

Figure 1: RHI as the decision between the common cause hypothesis ($C = 1$, left), i.e. all sensory input is caused by the rubber hand, and the separate causes hypothesis ($C = 2$, right), i.e. visual input is caused by the rubber hand and the proprioceptive and tactile input is caused by the real hand. X: position of hand, T: time points of brush strokes, χ_v: spatial visual input, τ_v: temporal visual input, χ_p: proprioceptive input, τ_t: tactile input. The image is from Samad, Chung, and Shams [13] and was released under a Creative Commons Attribution License.

Recently, an increasing number of scientists [2, 6, 19], have become convinced that humans maintain a causal model of their environment in their mind and constantly update it based on sensory information with a process that can be characterized in terms of Bayesian inference. Causal structures are a comparatively sparse representation of knowledge and can be used to make predictions about future observations and, crucially, the impact of interventional changes to the environment (such as design elements in VR applications) on the outcome.

The BCIBO model [13] explains the RHI as the participant's inference of a single cause of multisensory input, i.e. them reaching the conclusion that all sensory signals are caused by the stroking of the rubber hand. The model abstracts the sensory input of the brain down to two categories: *spatial* information which indicates the position of the rubber and/or real hand and *temporal* information which indicates the time points at which the brushes touch both (or one) of the hands. The latter has been shown to have an important influence on the illusion [1]: If the time sequences of the brush strokes are nearly identical (synchronous stimulation), the illusion often manifests in a majority of the participants. If the time sequences are out of synchronization (asynchronous stimulation), most participants do not experience the illusion.

Spatial and the temporal information comprise two sources of sensory information, respectively. The spatial information is provided by vision (χ_v) and proprioception (χ_p). χ_v can only provide information about the rubber hand (since the real hand is hidden from view) and χ_p only about the real hand. Temporal information is provided by vision (τ_v) and tactile signals (τ_t). Again, τ_v can only provide information about the rubber hand and τ_t only about the real hand.

The BCIBO model represents the hands' positions (χ_v and χ_p) in millimeters on a horizontal line relative to the body midline. It is assumed that the body and the table are roughly in parallel to each other. The timing of the brush stroke sequence (τ_v and τ_t) is represented by the time of the first brush stroke (in milliseconds) after the beginning of the trial. Assuming all of the brush strokes are separated by the same time interval (e.g. 1000 milliseconds), the time point of the first brush stroke provides enough information to represent the entire time series of strokes. Therefore, if τ_v, the time point for the real hand, is close to τ_t, the time point for the rubber hand, the brush stroking sequences are synchronized.

The inference about the causes of the sensory input is codified in the model as the decision between two hypotheses: the common cause (C_1) and the separate causes (C_2) hypothesis (s. Fig. 1). C_2 postulates the true state of affairs: χ_v and τ_v are caused by the rubber hand and χ_p and τ_t by the real hand. Hypothesis C_1 disowns the real hand and assumes the rubber hand to be the single cause of χ_v, τ_v, χ_p and τ_t. This causal attribution of the tactile and proprioceptive input to the rubber hand corresponds to two phenomena often reported by participants experiencing an RHI: touch referral and proprioceptive drift [17]. The latter is often reported by RHI participants and refers to a recalibration of the estimated position of the real hand based on proprioceptive input, in which said estimate shifts from the real hand towards the rubber hand. It should be noted though, that the drift often does not "reach" the rubber hand, as indicated by an average reported proprioceptive estimate that is often somewhere in between the real and the rubber hand.

We can arrive at the posterior probability of C_1 and C_2 by applying Bayes' Theorem:

$$p(C|\chi_v, \chi_p, \tau_v, \tau_t) = \frac{p(\chi_v, \chi_p, \tau_v, \tau_t|C)p(C)}{p(\chi_v, \chi_p, \tau_v, \tau_t)} \quad (1)$$

where C is a binary variable with $C = 1$ indicating the probability of the common cause and $C = 2$ indicating the probability of the separate causes hypothesis.

Samad et al. [14] have published a well-documented graphical user interface (GUI) for conducting Bayesian causal inference, written in the commercial programming language MATLAB. The GUI allows the user to easily set the parameter values of the model and displays a contour plot of the resulting posterior distribution. It contains a 2-dimensional Bayesian causal inference model, which is equivalent to the BCIBO model provided the user sets the parameters according to Samad, Chung, and Shams [13].

2.2 Critical Evaluation

Samad, Chung, and Shams [13] used Gaussians for all of the distributions in (1). This decision was probably made both for theoretical and practical reasons, since normal distributions allow for comparatively easy algebraic manipulation. They strove to choose "realistic values" (p. 6) for means and standard deviations and – for the most part – succeeded in this endeavor. For example, the standard deviation of τ_v and τ_t was set to 20 ms based on previous research [5].

X and T denote the priors for the position of the hands and the time point of the first brush stroke, respectively. Samad, Chung, and Shams [13] chose a "large number" (p. 6) as the standard deviation in order to approximate a uniform distribution. The exact value is not mentioned in the paper, but according to private correspondence [16] it was 10^{35}.

If a model attempts to approximate the data generating function of an aspect of human cognition, it needs to use realistic values for its parameters. 10^{35} is an unimaginably large number for humans and therefore it is implausible that such a number would be used in computations underlying the human psyche. To put the magnitude of this number into perspective: On the spatial dimension of the model, 10^{35} mm is around 1000 times larger than the length of the observable universe, and on the temporal dimension it is several levels of magnitude larger than the age of the universe. On top of this, a standard deviation covers only around 68% of a normal distribution, i.e. the values we could reasonably expect with this prior are even larger.

Samad, Chung, and Shams [13] ran the model for different distances between the real and rubber hand. Lloyd [11] has found that an increased distance between the real and rubber hand leads to a decrease in SoBO. Samad, Chung, and Shams [13] computed the posterior probability of $C = 1$ for different distances between the real and rubber hand and found results similar to Lloyd [11]. The

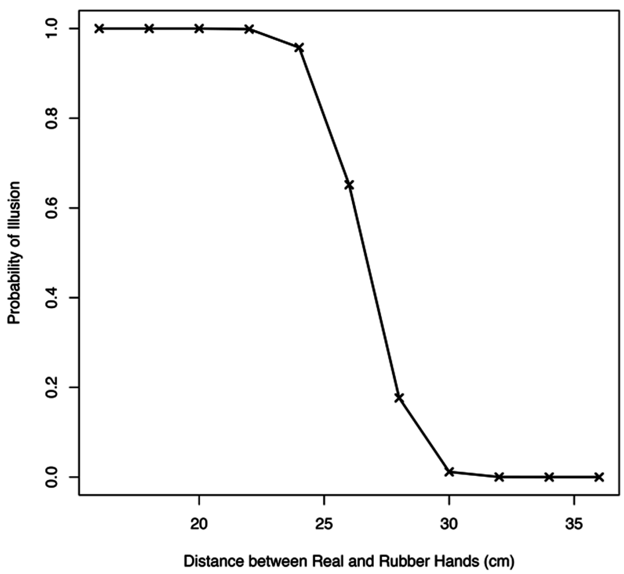

Figure 2: posterior probability of C_1 for different distances between real and rubber hand as presented by Samad, Chung, and Shams [13]. The image was released under a Creative Commons Attribution License.

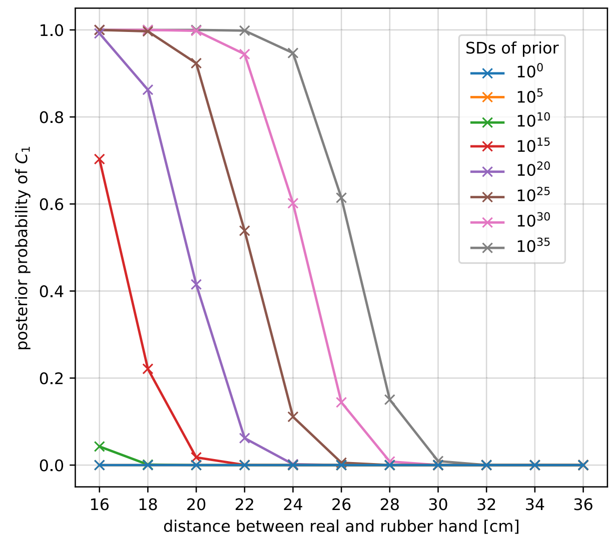

Figure 3: posterior probability of C_1 for different distances between real and rubber hand and across different magnitudes of standard deviations (SDs) for the prior

distance of 16 cm is equivalent to the placement of the real hand at 16 cm and the rubber hand at 32 cm away from the body's midline. Both positions are commonly found in RHI experiments [13]. An increase in the distance can be seen as moving the rubber hand further and further away from the real hand, eventually crossing the body's midline, similar to Lloyd's [11] experimental setup. The results of Samad, Chung, and Shams's [13] simulation are shown in Fig. 2.

We implemented the model ourselves and ran it for the same distances as Samad, Chung, and Shams [13] and across different standard deviations of the priors. Our code, written in the programming language Python [4, 8, 18], version 3.9.1, and licensed as free software under the MIT license, is available on GitLab [15]. As can be seen in Fig. 3, the results for the standard deviation used by Samad, Chung, and Shams, 10^{35}, closely resemble their results (cf. Fig. 2), indicating a successful re-implementation of the BCIBO model.

Fig. 3 shows that the posterior probability of C_1 for all distances declines with smaller choices of the prior's standard deviation. In order to be in line with empirical results [11], a good model of SoBO should predict high chances of a body ownership illusion (BOI) occurring for a 16 cm distance. However, for a standard deviation of 10^{10} mm the chance of experiencing a BOI at 16 cm is below 0.05. 10^{10} mm is equivalent to 10000 km, longer than the Great Wall of China, i.e. still a very implausible assumption for the position of one's hand in space.

Clearly a reasonable prior would allocate very little probability to hand positions outside of arm's reach. In our opinion, a sensible choice for the spatial prior's standard deviation would be 400 mm, because the 99% confidence interval of such a distribution spans the interval $[-931\text{mm}, 931\text{mm}]$, an area that is just large enough to cover many humans' arm spans. However, the BCIBO model fails with such reasonably narrow prior distributions, predicting microscopic posterior probabilities well below 0.001.

We therefore think that the model is in need of revision and will provide some speculations about how the model could be improved in the following section.

3 Future Research

Since choosing reasonably narrow normal distributions as priors yields unrealistic results, other choices of distributions have to be taken into consideration. One option that presents itself is a truncated normal distribution, i.e. a normal distribution that is cut off at the two ends of an interval. In case of the BCIBO model, this interval should correspond to the reach of one's arm. Input outside of said interval is not supported by the distribution and will therefore have zero probability. In psychological terms this could be understood as higher levels of cognitive processing flat out rejecting any processed visual or proprioceptive signals that suggest one's hand is outside of arm's reach. Ideally, the supported interval should be set to an individual's arm span to predict interindividual differences in the experience of BOIs.

Perhaps the most obvious choice for an alternative prior is a uniform distribution. After all, Samad, Chung, and Shams [13] only chose such a wide normal distribution as their prior in order to approximate a uniform distribution. Another idea would be to use a bimodal Gaussian mixture model: Each mode would be at the position of one of the shoulders, the implication being that (on the horizontal scale of the BCIBO model) one's hands are very likely to be close to the shoulders, likely to be near the body's midline and less likely to be in the periphery.

We are currenlty working on implementing some of the changes suggested above. In certain cases this can be done by algebraically arriving at a new equation for calculating the posterior. However, some posteriors can simply not be solved analytically. An alternative is to use Markov chain Monte Carlo sampling algorithms, which are a lot more flexible, but come with their own pitfalls, such as failure to converge.

After settling on a model with plausible parameters, a possible next step would be to see whether it can predict interindividual differences in empirical data. For example, one prediction of the model is that participants with higher proprioceptive acuity should have a lesser propensity to experience the illusion. Hence, a suggestion for future research would be to adjust the narrowness of the model's proprioceptive likelihood function according to a participant's proprioceptive acuity and see if this model component can predict differences in reported SoBO. VR should be the research paradigm of choice in such an experiment, because it allows for accurate assessment and manipulation of both the spatial and temporal information in the model through the recording of motion capture data and its (possibly manipulated) "playback" in VR.

4 APPLICATION

A possible application of the model would be to help designers of VR-related hardware to decide on tolerable levels of accuracy both for gathering and displaying spatial and temporal information. For example, a producer of head-mounted displays (HMDs) might have to decide between several design options all with different levels of accuracy and production costs. HMDs receive a time series of motion capture data as input and display them as a virtual environment. A well-working version of the BCIBO model would be able to predict the average user's SoBO based on the discrepancy between the *actual* motion capture positions and time points and the *virtual* positions and time points. Given the goal is to maximize the user's SoBO, the BCIBO model is able quantify the trade-off between the spatial and temporal inaccuracies of the system in terms of the probability of a BOI.

An application that presumably lies further in the future is the use of the BCIBO model as a component in a VR user model. A *user model* (e.g. Horvitz et al. [7]), as the name implies, tries to model relevant states of the user. An accurate SoBO user model could detect when a user's SoBO is slipping and enact counter-measures in the virtual environment. For example, in order to reinforce the SoBO a stimulus that encourages hand-based interaction could be presented. This would nudge the user towards looking at their virtual body which in turn should strengthen their SoBO.

5 CONCLUSION

In conclusion, while we think that the BCIBO model is a commendable step towards a computational explanation of SoBO we think that it is in need of revision due to its unrealistically wide prior distributions. It is our belief that a good model of SoBO will improve both our understanding of SoBO and the design of VR applications that rely on an embodied user experience.

ACKNOWLEDGMENTS

The authors wish to thank Majed Samad and Tim Schürmann. This work was supported by a grant from the German Research Foundation (DFG) to the IRTG 1901, "The Brain in Action".

REFERENCES

[1] M. Botvinick and J. Cohen. "Rubber Hands 'Feel' Touch That Eyes See". In: *Nature* 391 (1998), p. 756. DOI: 10.1038/35784.

[2] A. Clark. "Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science". In: *Behavioral and Brain Sciences* 36.03 (2013), pp. 181–204. DOI: 10.1017/S0140525X12000477.

[3] W. S. Geisler and D. Kersten. "Illusions, Perception and Bayes". In: *Nature Neuroscience* 5.6 (2002), pp. 508–510. DOI: 10.1038/nn0602-508.

[4] C. R. Harris et al. "Array Programming with NumPy". In: *Nature* 585.7825 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[5] I. J. Hirsh and C. E. Sherrick Jr. "Perceived Order in Different Sense Modalities." In: *Journal of experimental psychology* 62.5 (1961), pp. 423–432. DOI: 10.1037/h0045283.

[6] J. Hohwy. *The Predictive Mind*. First edition. Oxford, United Kingdom: Oxford University Press, 2013. 282 pp. ISBN: 978-0-19-968673-5.

[7] E. Horvitz et al. "The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users". In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. UAI'98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 256–265. ISBN: 978-1-55860-555-8.

[8] J. D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[9] K. P. Körding et al. "Causal Inference in Multisensory Perception". In: *PLOS ONE* 2.9 (2007), pp. 1–10. DOI: 10.1371/journal.pone.0000943.

[10] E. Lewis and D. M. Lloyd. "Embodied Experience: A First-Person Investigation of the Rubber Hand Illusion". In: *Phenomenology and the Cognitive Sciences* 9.3 (2010), pp. 317–339. DOI: 10.1007/s11097-010-9154-2.

[11] D. M. Lloyd. "Spatial Limits on Referred Touch to an Alien Limb May Reflect Boundaries of Visuo-Tactile Peripersonal Space Surrounding the Hand". In: *Brain and Cognition* 64.1 (2007), pp. 104–109. DOI: 10.1016/j.bandc.2006.09.013.

[12] J. F. Magnotti, W. J. Ma, and M. S. Beauchamp. "Causal Inference of Asynchronous Audiovisual Speech". In: *Frontiers in Psychology* 4 (2013), pp. 1–10. DOI: 10.3389/fpsyg.2013.00798.

[13] M. Samad, A. J. Chung, and L. Shams. "Perception of Body Ownership Is Driven by Bayesian Sensory Inference". In: *PLOS ONE* 10.2 (2015). Ed. by M. A. Lebedev, pp. 1–23. DOI: 10.1371/journal.pone.0117178.

[14] M. Samad et al. *Bayesian Causal Inference Toolbox*. 2017. URL: https://github.com/multisensoryperceptionlab/BCIT (visited on 02/05/2021).

[15] M. Schubert. *The Bayesian Causal Inference of Body Ownership Model*. 2021. URL: https://gitlab.com/moritz_schubert/bcibo (visited on 02/04/2021).

[16] T. Schürmann. *Implementation Body-Ownership-Modell von Samad et al.* E-mail. Mar. 23, 2020.

[17] M. Tsakiris. "The Sense of Body Ownership". In: *The Oxford Handbook of the Self*. Oxford Handbooks. 2011, pp. 180–203. ISBN: 978-0-19-954801-9.

[18] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature methods* 17.3 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[19] W. Wiese and T. K. Metzinger. "Vanilla PP for Philosophers: A Primer on Predictive Processing". In: *Philosophy and Predictive Processing*. Ed. by T. K. Metzinger and W. Wiese. Frankfurt am Main: MIND Group, 2017, pp. 1–18. ISBN: 978-3-95857-302-4. DOI: 10.15502/9783958573024.